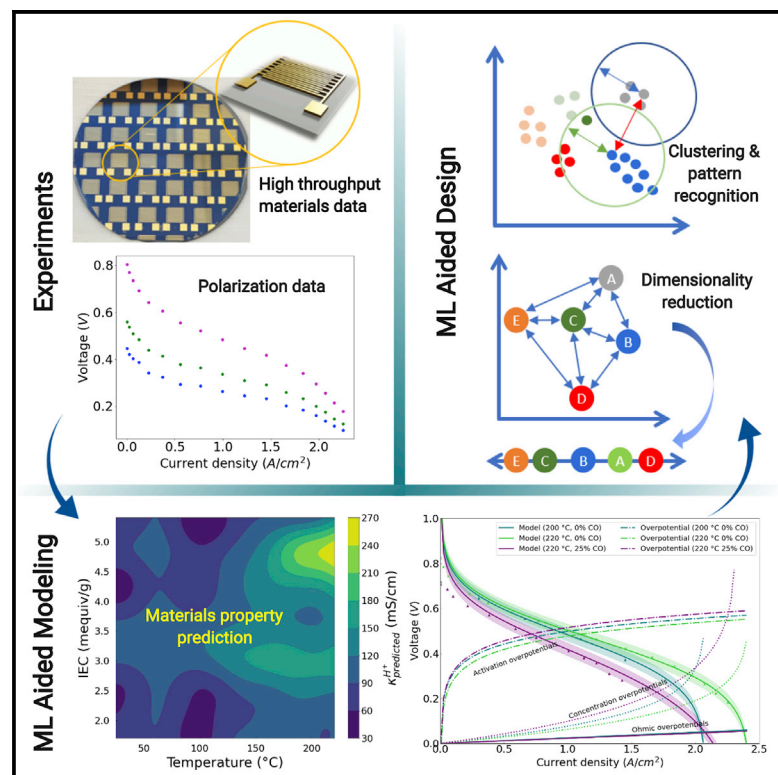# Patterns

Machine learning for guiding high-temperature PEM fuel cells with greater power density

Graphical abstract

Highlights

- A framework that integrates materials data with device operation via ML

- ML of synthetic data from an experimentally verified current distribution model

- Dimension reduction and clustering identify pathways for improved HT-PEMFCs

Authors

Luis A. Briceno-Mena,
Gokul Venugopalan,
José A. Romagnoli,
Christopher G. Arges

Correspondence

jose@lsu.edu (J.A.R.),
carges@lsu.edu (C.G.A.)

In Brief

A multilevel modeling and data analysis framework was constructed for high-temperature polymer electrolyte membrane fuel cells (HT-PEMFCs) and their electrolyte materials. The framework used Machine Learning tools (e.g., support vector regression, dimension reduction, and clustering) that seamlessly linked materials characteristics with fuel cell device performance and design, allowing for the accelerated discovery of material property attributes and fuel cell operating parameters and configurations that achieve greater power density and efficiency while co-currently addressing costs.

CellPress

Article

Machine learning for guiding high-temperature PEM fuel cells with greater power density

Luis A. Briceno-Mena,[1] Gokul Venugopalan,[1] José A. Romagnoli,[1,*] and Christopher G. Arges[1,2,*]

[1]Cain Department of Chemical Engineering, Louisiana State University, Baton Rouge, LA 70803, USA
[2]Lead contact
*Correspondence: jose@lsu.edu (J.A.R.), carges@lsu.edu (C.G.A.)
https://doi.org/10.1016/j.patter.2020.100187

THE BIGGER PICTURE Renewable energy and energy efficiency are crucial for achieving global sustainability goals. In this context, there is need for the development of new materials that realize high-performing and low-cost power sources. At the same time, advances in computational power, simulation, and Machine Learning enable researchers to explore large amounts of data, providing inspiration and tools for the design of new systems. In this work, we combined experiments with modeling and data analysis tools to build a framework for the study and development of high-temperature polymer electrolyte membrane fuel cells (HT-PEMFCs). The framework used Machine Learning tools (e.g., support vector regression, dimension reduction, and clustering) that seamlessly linked materials characteristics with fuel cell performance. This allowed for the accelerated discovery of material properties and fuel cell operating parameters that achieve greater power density while co-currently addressing costs.

1 2 3 4 5   **Proof-of-Concept:** Data science output has been formulated, implemented, and tested for one domain/problem

---

SUMMARY

High-temperature polymer electrolyte membrane fuel cells (HT-PEMFCs) are enticing energy conversion technologies because they use low-cost hydrogen generated from methane and have simple water and heat management. However, proliferation of this technology requires improvement in power density. Here, we show that Machine Learning (ML) tools can help guide activities for improving HT-PEMFC power density because these tools quickly and efficiently explore large search spaces. The ML scheme relied on a 0-D, semi-empirical model of HT-PEMFC polarization behavior and a data analysis framework. Existing datasets underwent support vector regression analysis using a radial basis function kernel. In addition, the 0-D, semi-empirical HT-PEMFC model was substantiated by polarization data, and synthetic data generated from this model was subject to dimension reduction and density-based clustering. From these analyses, pathways were revealed to surpass 1 W cm^{-2} in HT-PEMFCs with oxygen as the oxidant and CO containing hydrogen.

INTRODUCTION

High-temperature polymer electrolyte membrane fuel cells (HT-PEMFCs), which operate in the temperature range of 120–250°C, offer simpler heat and water management when compared with today's conventional low-temperature PEMFC variants.[1,2] The elevated temperature operation of HT-PEMFCs reduces the size of the fuel cell stack radiator, while also eliminating any need for feed gas humidification. Despite the advantages of HT-PEMFCs, they have been maligned over time for

vehicle applications due to their low power density, high platinum group metal loadings,[3] unsatisfactory stability,[1,4,5] and limited temperature range and water tolerance.[6,7] Because of these limitations, this fuel cell platform has largely been relegated to stationary power and niche applications.[6]

In 2016, Los Alamos National Laboratory reported[7] a superior polymer electrolyte membrane (PEM) and ionomer binder for HT-PEMFCs based on a phosphoric acid (H$_3$PO$_4$) imbibed polycation. Unlike the standard bearer H$_3$PO$_4$ containing polybenzimidazole (PBI) variant[2,6,8,9] (and other similar to tertiary

amine containing polymers[10]), the electrostatic interactions between the tethered cation moiety and phosphate anion in H_3PO_4 imbibed polycations permitted the use of wider temperature and humidity ranges without worry of H_3PO_4 evaporation and leaching.[7] Venugopalan et al.[11] improved on this concept through the blending of PBI with the polycation to facilitate more H_3PO_4 uptake and greater ionic conductivity (area specific resistance [ASR] < 0.02 Ω-cm^2). The membrane blend was stable at 220°C and had tolerance to 40% relative humidity at 80°C while showing remarkably high ionic conductivity over the temperature range of −40 to 240°C.[11,12] Notably, both groups attained promising power density values near 0.7 W cm^{-2} with pure oxygen as the oxidant.[7,10] The advent of the ion-pair polymer electrolytes has renewed interest in HT-PEMFCs for vehicular applications. But, success for this technology still necessitates improved power density, especially with air as the oxidant, while also reducing platinum group metal loadings (e.g., <6 g$_{Pt}$/vehicle; rated at 90 kW$_{net}$ power[13]).

Perhaps the most significant barrier to better HT-PEMFC power density is the presence of liquid H_3PO_4 present in the electrode layers with ionomer binder. The phosphate anions in H_3PO_4 can adsorb to the platinum electrocatalyst surface, blocking reactant sites, and thus causing large activation overpotentials.[14] Furthermore, the presence of liquid acid and hydrocarbon binder stymie delivery of gas reactants to the electrocatalyst surface leading to large concentration overpotentials.[15] These sources of overpotential significantly hamper the power density of HT-PEMFCs and need to be addressed if this platform will ever emerge as a serious competitor for powering light-duty and heavy-duty vehicles (LDVs and HDVs).

Materials science and engineering has played a central role in the evolution of improved electrochemical technologies like fuel cells and batteries. However, the design of new materials is often "Edisonian" and the timeline from discovery to implementation in commercial devices is costly and far too long. The emergence of Machine Learning (ML) methods across a variety of disciplines (e.g., chemical processes and biology), combined with the ubiquity of powerful computers, has motivated efforts in using these tools for streamlining the timeline from materials discovery to commercialized products.[16] In the context of electrochemical energy storage and conversion, ML has been adapted to capture molecular interactions of materials and relating them to bulk properties so that potential candidates for components and devices can be quickly identified.[17–20] It is also effective for predicting device performance from experimental operational data.[21] More recently,[22,23] the implementation of ML approaches coupled with single components modeling is being explored as way to bridge the gap between materials property predictions and device performance. However, such strategies have not been widely used for PEM-based electrochemical systems (e.g., fuel cells and electrolyzers).

In this work, ML and data-driven analysis tools were leveraged to identify HT-PEM material properties for achieving high-power HT-PEMFCs (~1 W cm^{-2} with oxygen as the oxidant). The approach first starts with a physics 0-D model that is shown to predict HT-PEMFC polarization behavior data with a few parameters estimated using a Jaya optimization algorithm. Support vector regression using a radial basis kernel was performed to relate material attributes, such as ion-exchange capacity (IEC), to material properties (H_3PO_4 uptake and ionic conductivity).

This activity informed the ohmic overpotential and parts of the concentration overpotential contributions in the 0-D model that describes HT-PEMFC polarization. This connected the materials property predictions to device performance. Verifying the model with existing datasets instilled confidence for generating data from the model for further ML activities using density-based clustering via Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) and dimensionality reduction via Uniform Manifold Approximation and Projection (UMAP). These activities revealed that in order to surpass the 1 W cm^{-2} power density objective with air as the oxidant, the oxygen diffusivity coefficient in the cathode would need to improve by 2.2x with respect to the base case (H_3PO_4 imbibed polycation),[11] while maintaining facile proton conductivity (~0.2 S cm^{-1}) in the electrode layers. Achieving such a breakthrough could be achieved by new permeable and highly conductive cathode ionomer binders and by applying 187 kPa$_{abs}$ of back pressure on the cell. It is worth mentioning that this pathway for a high-power HT-PEMFC necessitated 218°C operating temperature, which is possible with acid-imbibed polycation materials but not plausible with the standard bearer phosphoric acid containing polybenzimidazole materials. Further, the simulated HT-PEMFC with ~1 W cm^{-2} operated with a hydrogen fuel stream containing 12% CO and had a platinum catalyst loading in the electrodes as low as 0.23 mg$_{Pt}$ cm^{-2}. These attributes are particularly important for addressing fuel cell capital costs, large-scale manufacturing, and running the fuel cell on low-cost hydrogen derived from steam reformed methane. In summary, ML and modeling at various scales (e.g., materials properties to device performance) was developed for HT-PEMFCs and was shown to be effective with existing datasets and for identifying pathways to achieve high power density.

RESULTS

For a modeling and data analysis framework to enable the study of materials properties and its effects on the fuel cell performance, it must be implemented in a suitable platform in which the materials models can be embedded. Also, it is desirable that the model implementation allow for the generation of large datasets so that a great number of possible configurations can be explored, with low computational cost. Hence, in this work, a Python-based, zero-dimensional, steady-state, continuum approach was adopted. The model incorporates the influence of 10 input variables (hydrogen stoichiometric ratio, S_{H_2}; oxygen stoichiometric ratio, S_{O_2}; temperature, T; pressure, P; membrane IEC, IEC_{mem}; ionomer binder IEC, IEC_{io}; membrane thickness, δ_{mem}; ionomer binder thickness, δ_{io}; carbon monoxide to hydrogen ratio in feed stream, CO/H_2; and platinum loading, L_c) as well as current density (i) to predict the voltage of the fuel cell. Seven fitting parameters were included in the model: reference exchange current density for the anode and the cathode, $i_{0-ref}^{an,cat}$; an empiric parameter for the limiting current at the cathode and the anode side, $K_{an,cat}$; the charge transfer coefficient for the cathode and the anode, $\alpha_{an,cat}$; an empiric parameter for the effect of phosphoric acid on mass transfer, γ; and an empiric parameter for the concentration overpotential, B. The details for the model are shown in the Supplemental information. This model served as a bridge between the materials
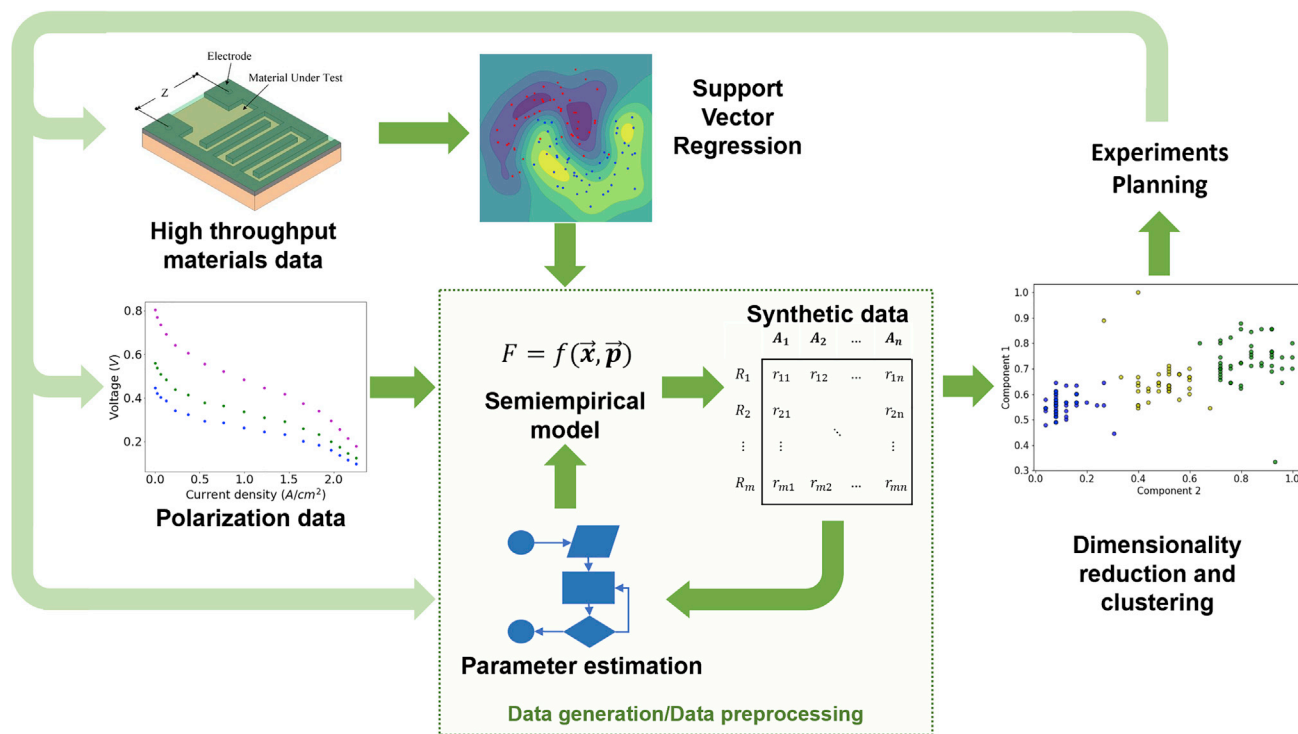
Figure 1. Schematic representation of the modeling and data analysis approach applied to study the use of ML methods for the fast development of new fuel cell designs

experimental data, ML predictive models, and the system-level data analysis so identified patterns from ML could be related to the materials properties in addition to fuel cell configurations and operating conditions. The results from these analyses give a set of clusters that were associated with HT-PEMFC power density. Furthermore, the distribution of the clusters can be connected to physically meaningful variables, thus informing new experiments and possible designs.

Here, we selected HDBSCAN clustering preceded by UMAP for dimension reduction to find patterns (i.e., clusters) in the data associated with high HT-PEMFC power density values and to identify the more contributing variables to the clustering. We implemented HDBSCAN with 35 *min_samples*, and UMAP with 10 *n_neighors*, 0.1 *min_dist*, and 3 *n_components*.[24–26] The resulting clusters were further analyzed using the subspace greedy search (SGS) algorithm described by Zhu et al.[27] The variables identified with the largest contributions to the separation between the classes are the same as those with the higher first-order sensitivity indices from the Sobol's global sensitivity analysis, as described in the Supplemental information. Furthermore, a set of candidate combinations for improving the peak power density of the fuel cell was found from the cluster analysis.

Figure 1 presents a scheme of the modeling and ML approach for identifying electrochemical material property attributes for improving HT-PEMFC power density. A salient aspect of this figure is the methods of learning (i.e., a circular loop) that feed into the 0-D, semi-empirical model that describes single-cell HT-PEMFC polarization behavior. The utility of a 0-D model is 2-fold: First, having a model enables the generation of synthetic data when experimental data are scarce, thus providing a way to

develop and test design and optimization approaches simultaneously while more data are generated. When experimental data are available, the block marked in Figure 1 (data generation/data preprocessing) can be substituted by only data preprocessing. Second, traditional optimization approaches that require a model can be compared with ML data-driven approaches that do not need a model to be implemented. The overall product of this combinatorial approach is a flexible framework that can exploit the benefits of ML techniques. For instance, the ability to visualize the search space allows for human intuition to be used for the analysis so that less than optimal options, which may be overlooked by an optimization algorithm, can be identified and explored. Also, the weight of the model assumptions can be balanced against the cost of producing new experimental data, and the resolution and range needed for such experimental data can be identified a priori. Other phenomena, such as gas transport in electrode layers, can be detailed in a similar fashion as presented here for ionic conductivity with support vector regression (SVR) and then be incorporated into the analysis without having to perform additional polarization experiments. In addition to all of these strategies, the ML methods enable identification and filtering of noise in the data fed to the algorithms, thus compensating some of the drawbacks of a simplified semi-empirical model. It is envisioned that the proposed framework fosters model development as well as identifying alternative fuel cell operating parameters and materials that boost performance.

The 0-D, semi-empirical model was built off existing HT-PEMFC models that are physics-informed with distributed parameters.[28–30] But, the model also features some semi-empirical
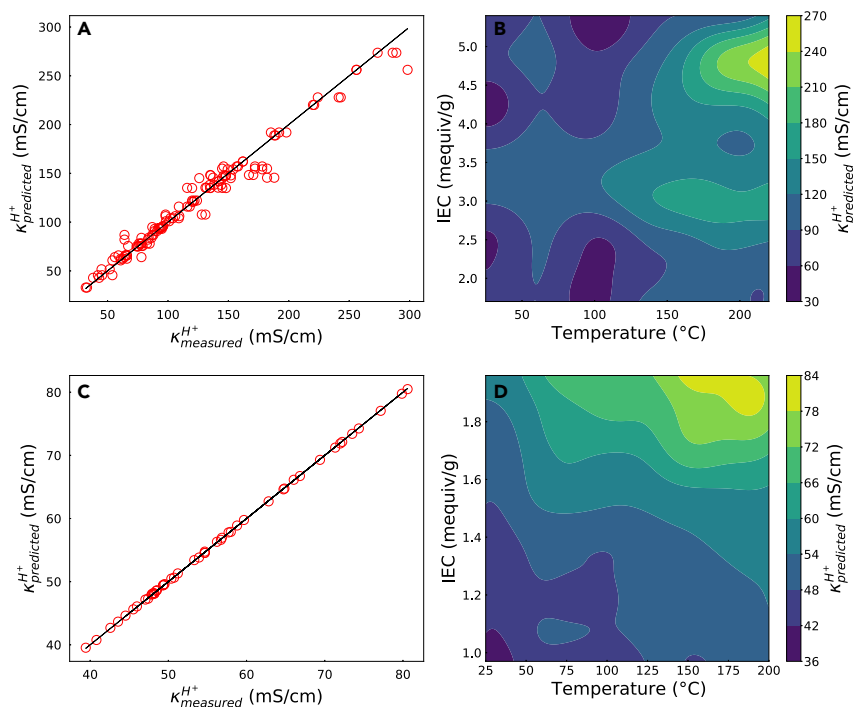
Figure 2. Model for ionic conductivity as a function of temperature and IEC using a support vector regression with a radial basis function kernel

(A–D) (A and B) PEM and (C and D) thin ionomer films that are similar in thickness to the ionomer binder in electrode layers.

can be applied to a wide range of datasets. Hence, SVRM was selected for the prediction of bulk material properties (e.g., ionic conductivity) from material characteristics (e.g., IEC) under different operating conditions (e.g., temperature) because data were available for bulk HT-PEMs and H_3PO_4 imbibed polycation thin films. Although the Nernst-Planck relationship captures these parameters on ionic conductivity for ideal liquid electrolytes,[36] PEMs feature non-ideal behavior and complexity associated with their solid-state structure[37] and concentrated ionic groups. The SVRM was useful for determining the unknown underlying relationship between IEC and temperature and ionic conductivity in the acid-imbibed polymer electrolytes.

expressions[31–33] to describe concentration polarization overpotential losses that hail from transport phenomena in complex porous structures. The lumped model was selected because it accurately predicts data with few adjustable parameters, and it has relevant descriptors that factor how H_3PO_4 content in the membrane and CO in the feed gas stream impact HT-PEMFC polarization. The key assumptions for the model were continuum level, steady-state operation, ideal gas behavior, no gas crossover, no water transport considerations, and uniform distribution of catalyst and ionomer binder and gas transport properties across the electrodes and gas diffusion layers. The Supplemental information presents the model equations that include mass balances and reaction rate expressions, parameters for these equations, and descriptions for individual overpotential terms that govern HT-PEMFC polarization. Notably, a semi-empirical expression was deployed to account for the ionomer binder gas transport (i.e., diffusivity) and conductivity properties.
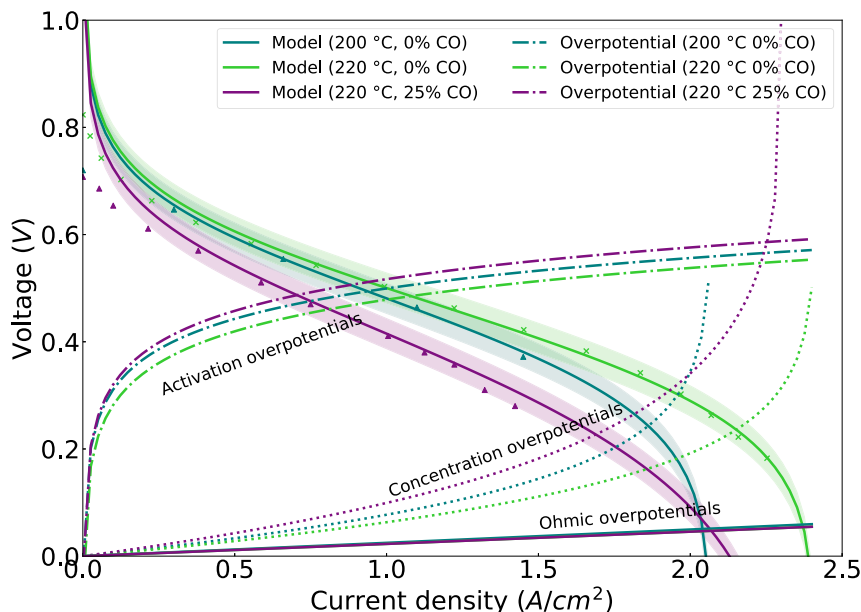
Machine learning embedded materials modeling

A wide variety of ML methods and implementations exist for data analysis. The selection of the appropriate method must consider aspects like the amount of data available, the purpose of the implementation, and the computational resources. A notable method is the so-called SVR Machines or SVRMs. The idea of SVRM was first introduced by Drucker et al.[34] as an application of the concept of support vector machines (SVMs) first proposed by Boser et al.[35] An SVM is an algorithm that looks for the hyperplane that separates two linearly separable data classes maximizing the distance between the hyperplane and the classes. If the data are not linearly separable, then a transformation, known as kernel function, is introduced to map the data into a new space in which it is linearly separable. This feature makes SVMs particularly flexible, as they

To demonstrate the feasibility of SVMs for relating modeling material property data that would feed into a device-level model, the approach created by Chang and Ling[38] was adopted. The Supplemental information provides a detailed description of the method and thin film ionomer conductivity data measured on interdigitated electrodes (IDEs), and IEC values of the high-temperature polymer electrolyte materials used as thin films and electrode binders (Tables S4 and S5). Bulk HT-PEM ionic conductivity and IEC data were taken from our previous publication.[11] Figure 2 conveys that the support vector machine models give an accurate prediction for the ionic conductivity as a function of material IEC (non–acid-imbibed values) and environmental temperature for both the PEM and the ionomer binder (characterized as a thin film). Figure S4 shows a similar plot except the IEC values are expressed as mmol of H_3PO_4 per mass. These models, which were informed by data, are embedded into the 0-D, semi-empirical HT-PEMFC device model, so that the polarization behavior of the fuel cell can be predicted accurately from the materials' characteristics.

It is recognized that a limited demonstration was only performed with support vector machine models using material IEC and ionic conductivity. Other materials, and their attributes, are also important, for example, electrocatalysts' electrochemically active surface area and reactivity (e.g., exchange current and symmetry factor) and reactant gas diffusion in catalyst layers. Our future work will look to factor these parameters and considerations into the polarization model. Attaining reaction kinetic and gas transport data will reduce the number of estimated parameters in the fuel cell model and will assist in realizing a more robust and rigorous prediction of HT-PEMFC polarization behavior. Here, ionic conductivity of the electrode binder and

HT-PEM were examined as a demonstration of the SVRM effectiveness and the availability of data.

Model validation

To assist ML and model validation, both material property data and HT-PEMFC polarization data were included. Material attributes and property data were analyzed using an SVRM in the previous section. The model parameters were estimated using the Jaya optimization algorithm.[39] Figure 3 presents the polarization curve from the model predictions and the polarization data (note: data are from beginning of life for the fuel cell). Table S6 shows the membrane electrode assembly (MEA) properties and gas reactant flow rates for polarization data presented in Figure 3. Table S7 presents the predicted parameters from the Jaya parameter estimation method for the HT-PEMFC model that successfully described the experimental polarization behavior from our previous work.[11] It is worth mentioning that this is the first time HT-PEMFC performance has been modeled and substantiated with H_3PO_4-imbibed polycations and polycation/PBI blends. These materials enable operation at temperatures above 180°C for the HT-PEMFC, which currently cannot be operated or considered with H_3PO_4 containing polybenzimidazole (PBI). The incorporation of polycations into the membrane and electrode anchors the phosphoric acid through electrostatic interactions between the phosphate anions and quaternary pyridinium groups. These electrostatic interactions stabilize the phosphoric acid in the membranes under challenging conditions (e.g., 220°C/0% relative humidity [RH] and 80°C/40% RH). Phosphoric acid doped PBI loses acid under these conditions.[3] Hence, the water produced in the fuel cell does not adversely affect the membrane and electrode ionomer stability (i.e., acid loss that causes a drop in conductivity) at the temperatures being considered because of the selection of the materials.

Figure 3 also shows the individual overpotential terms as a function of current density from the HT-PEMFC model. These overpotential terms govern the polarization of the HT-PEMFC. Conveying the individual overpotential terms against the HT-

PEMFC polarization curve makes it possible to pinpoint the greatest sources of resistance that hamper fuel cell power density. As expected, the activation overpotentials, which arise from electrode kinetics, mainly the sluggish oxygen reduction reaction, cause the largest source of polarization in the cell when extracting up to ~2 A cm^{-2} of current density (see Figure S5). The ohmic overpotential, which arises from the membrane and contact resistance, is very low but agrees with other reports based on LT-PEMFCs.[40,41,42] The low ASR observed from fuel cell experiments hailed from the HT-PEM's high ionic conductivity. The MEA had an ASR of 0.08 Ω-cm^2 at 100°C to 0.01 Ω-cm^2 at 220°C with non-humidified H_2/O_2.[11] This low ASR leads to a low ohmic overpotential across the different current density values. Operating the cell at higher temperatures (220°C as opposed to 200°C) promoted reaction kinetics and gas diffusion in electrode layers resulting in a reduction in activation and concentration overpotentials. As such, the cell displayed less polarization and a 58% improvement in peak power density.

The seven model parameters (see Table S7) to describe HT-PEMFC polarization were estimated using the Jaya optimization algorithm. The initial values were those reported from the literature for the previously reported data and defined by inspection for the case of K_{cat}. As a population-based heuristic parameter estimation method, Jaya will yield different results for various runs. Therefore, estimation of the parameter in the model required running the algorithm for 10 iterations and 100 runs to attain a measure of the variability of the algorithm (see Figure S3). The standard error for the estimation was 0.018 V, which, when combined with the convergence and stability of the parameter estimation algorithm, provides strong confidence on the model performance. We also used Sobol's global sensitivity analysis to identify which parameters were the most influential on the variability of the model (see Figures S1 and S2). These results provide information about the physical significance of the parameters involved in the model. The charge transfer coefficient for both the cathode and the anode, as well as the empirical parameter for the acid uptake in the concentration overpotential, were shown to be the most influential on polarization behavior in the low current density regimen.

Design space exploration using machine learning-based visualization

When selecting an appropriate ML implementation to inform the design of physical system, several considerations must be made.
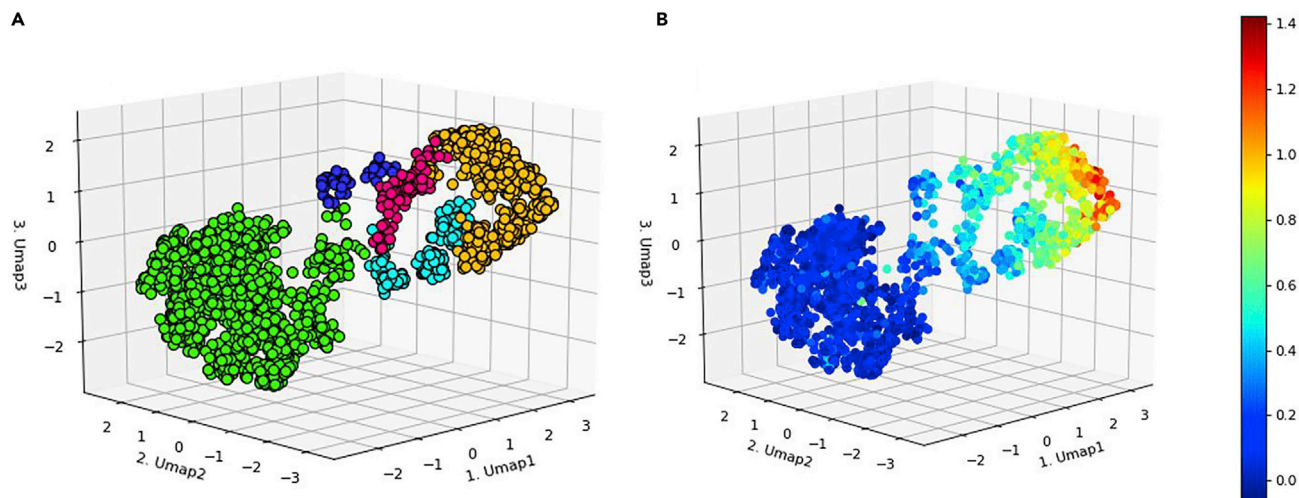
Figure 4. Clustering visualization for the synthetic data
(A) Clusters generated by HDBSCAN with preprocessing using UMAP (cluster 1 in green; cluster 2 in cyan; cluster 3 in pink; cluster 4 in blue; cluster 5 in yellow).
(B) Clusters with power density superimposed as color map.

First, due to the explorative nature of the task, no assumptions can be made regarding the shape of the clusters. In addition to this, the lack of a priori knowledge about the data and how it should look makes it difficult to establish the number of clusters needed. To deal with these issues, a density-based clustering (DBC) approach was adopted. DBC finds clusters of any shape, as opposed to centroid-based clustering, which assumes a shape (i.e., a sphere of equal variance) for the clusters. Also, in DBC, data points that are in sparse regions are not required to be assigned to a cluster and are instead identified as noise. Therefore, DBC gives us two advantages: (1) we remove the spherical equal variance assumption, allowing for different shapes in our clusters, and (2) we are now able to detect and deal with noise. However, the resolution parameter (namely the radius of the region that is being considered as dense or sparse) is hard to get, especially in higher dimensional data that we cannot easily visualize. This challenge can be overcome by introducing a hierarchical component. Hierarchical clustering finds nested relationships among data instead of flat partitions in the data. This characteristic allows for a more complex clustering with higher resolution by further diving clusters until a stopping criterion is met. Thus, combining HC and DB we get a clustering algorithm that makes no assumptions about the shape of the clusters and uses hierarchical clustering to improve the resolution eliminating the need of finding a correct resolution parameter.

In HDBSCAN, as introduced by Campello et al.,[24] the density is estimated by finding the radius of the region for which the said region is considered dense using mutual reachability as distance measure. The stopping criterion for the hierarchical approach is how long a cluster persists in time before it splits into two clusters or vanishes into noise. The computational cost is, however, $O(n^2)$, which means it does not scale well. In this work, we use the implementation proposed by McInnes and Healy[25] in which Dual Three Boruvka's for Euclidian Minimum Spanning Trees algorithm is used to enhance performance, bringing the computational cost to $O(N\log N)$.

Since HDBSCAN in based on density, we can expect that the performance would decrease with highly sparse data. Also, the

HDBSCAN performance suffers with high-dimensional data. Therefore, using a dimension reduction, preprocessing could enhance the clustering output. UMAP is a neighbor graphs-based dimension reduction technique, meaning that a graph is first built for the high-dimensional data and then it is embedded in a low-dimensional space using a force-directed layout using cross entropy to measure the distance between the high-dimensional graph and the low-dimensional graph. The resulting low-dimensional embedding is denser, providing a better starting point for HDBSCAN.[26]

With an experimentally substantiated semi-empirical HT-PEMFC model for ion-pair polymer electrolytes, it was possible to explore different configurations using a stochastic generator. To produce such configurations, 10 input variables (hydrogen stoichiometric ratio, S_{H2}; oxygen stoichiometric ratio, S_{O2}; temperature, T; pressure, P; membrane ion-exchange capacity, IEC_{mem}; ionomer binder ion exchange capacity, IEC_{io}; membrane thickness, δ_{mem}; ionomer binder thickness, δ_{io}; carbon monoxide to hydrogen ratio in feed stream, CO/H_2; and platinum loading, L_c) were allowed to take different values around the base case (see Table S1) in a uniform distribution within a given interval (e.g., ± base value×40%). For example, a given configurations would include a lower catalyst loading, L_c, and higher temperature and pressure, while another configuration would have the same catalyst loading but with a smaller membrane thickness, δ_{mem}. Gaussian noise was introduced into the model calculations to account for the variability of the model predictions induced by its sensitivity to the parameter estimation, taking advantage of the fact that noise-filtering feature of HDBSCAN provides robustness to the model variability. The learning agent was then exposed to a variety of scenarios from the generated simulations from which it can then draw and identify patterns and then evaluate the likelihood of new configurations that performs well. The clusters of data can then be associated with desired objectives, such as higher values of peak power density, and then the variables that exert greater influence on the distance between clusters can be revealed and guide the factors and material properties that lead to improved HT-PEMFC power density. To illustrate this concept, Figure 4 shows the visualization for the
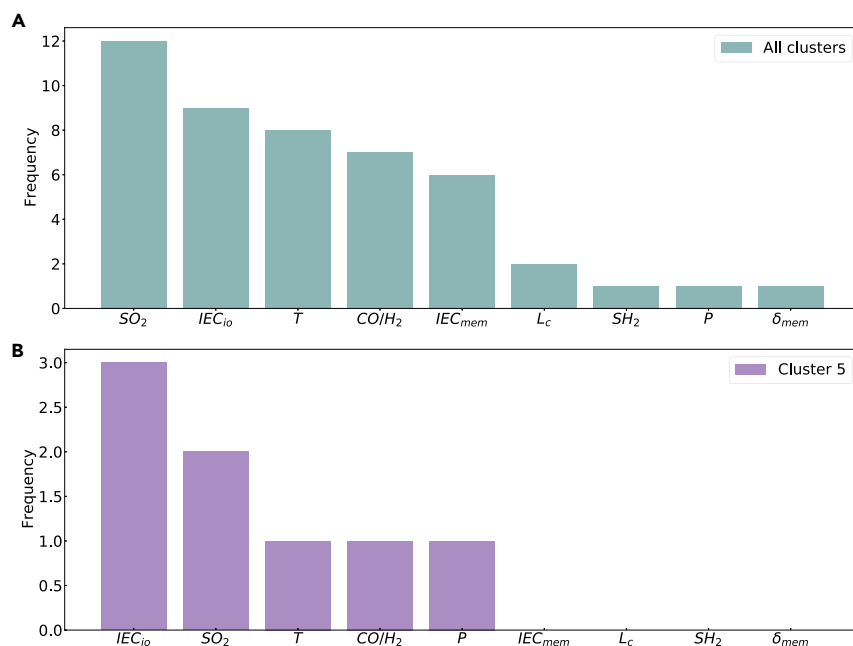
Figure 5. Frequency of subspaces appearing as main contributors to cluster formation as determined by the SGS analysis

(A) Frequency for the contributions for all the clusters.

(B) Frequency for the contributions for cluster 5.

clustering analysis using HDBSCAN with UMAP as preprocessing. Three major regions are seen in the cluster visualization: a region of low power density (cluster 1 in Figure 4A, and deep blue in Figure 4B), a high power density region in cluster 5, and a transition region in-between. Cluster 5 contains configurations that achieve higher power density values.

Although the combination of UMAP with HDBSCAN preserves the relationship of the high-dimensional space in the low-dimensional embedding, the physical interpretation of the low-dimensional variables Umap1, Umap2, and Umap3 is not readily accessible. Therefore, we applied a SGS algorithm analysis (see Figure 5 and Tables S2 and S3) in which the variables S_{O2}, T, and IEC_{io} were identified as the most influential in the formation of the clusters. The variables identified by the clustering and SGS analyses as the more influential on the peak power density are the same as those with the higher first-order index in the modified Sobol's global sensitivity analysis, which serves as cross validation for the UMAP-HDBSCAN clustering and SGS analysis approach. With this, not only are a set of candidate configurations (e.g., values for design and operating variables) attained, but it also provides insights as to which variables should be prioritized for enhancing HT-PEM performance and efficiency.

DISCUSSION

The ML-based modeling and analysis framework presented here enables fast identification of materials properties and device operating parameters that enhance HT-PEMFC performance. It is also worth noting that this approach could work even in the absence of any model since the clustering analysis can be performed using only experimental data. The hierarchical structure of the modeling strategy gives the modeler a flexible and powerful tool, in which the ML model is informed from basic physical knowledge of the HT-PEMFC. The model is also conducive for upgrading with additional descriptors as

more data are generated from experiments and/or made available in the literature and publicly accessible databases. Here, the data were primarily confined to the ionic conductivity and IEC of the HT-PEM and electrode binders and HT-PEMFC polarization. Future work will look to incorporate electrode reactivity and reactant/product species transport rates. In addition, this work deployed DBC through the HDBSCAN combined with UMAP implementation, and these tools allowed for the extraction of the underlying distribution of the data, that is, an accurate description of the polarization behavior of the system. From the cluster analysis, cluster 5 provides more combinations that are likely to surpass the 1 W cm^{-2} goal with oxygen as the oxidant. One of such combinations features a 2.2x increase in the oxygen diffusivity coefficient with respect to the base case, a back pressure of 187 kPa, a cell temperature of 218°C, and a proton conductivity value of ~0.2 S cm^{-1} in the electrode layers. Increasing the oxygen diffusivity in the electrode layer is paramount for reducing concentration overpotential and improving cell power density when using air as the oxidant. Although some of the candidate combinations will be deemed as unfeasible because of practicality concerns, high cost, or materials with desired properties cannot be realized, the search space for new configurations was significantly reduced, streamlining the amount of experimental permutations to be performed; and thus, accelerating the development of HT-PEMFC technology.

EXPERIMENTAL PROCEDURES

Resource availability
Lead contact
Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Christopher G. Arges (carges@lsu.edu).
Materials availability
This study did not generate new unique reagents.
Data and code availability
The datasets and code generated during this study are available at GitHub: https://github.com/lbrice1/FCSDAT.git.

Materials synthesis and characterization
The Supplemental information provides the details about the materials synthesis for QPPSf. The data for the bulk HT-PEM ionic conductivity and corresponding H_3PO_4 uptake were taken from a previous publication by Venugopalan et al.[11] The HT-PEMFC polarization data were also taken from the same publication.

The ML and data analyses performed in this work were augmented with new experimental data that examined thin film ionic conductivity of

H$_3$PO$_4$-imbibed polycations (i.e., quaternary benzyl pyridinium poly[arylene ether sulfone] [QPPSf]). QPPSf of varying IECs was synthesized based as described in the literature.[43,44] The Supplemental information provides a succinct synthesis procedure. IDEs used for measuring thin film ionic conductivity were manufactured using the procedure by Arges et al.[45,46] Thin films of QPPSf were deposited on the IDEs by spin coating a 1 wt% solution of QPPSf dissolved in n-methyl-2-pyrrolidone (NMP) at 4000 rpm for 45 s. The IDEs with thin film QPPSf were then heated at 120°C in nitrogen atmosphere to remove the excess NMP from the samples. H$_3$PO$_4$ was imbibed into the thin films by placing a drop of 85 wt% H$_3$PO$_4$ on the thin films for 10 min. The excess acid was removed by blot drying the IDEs carefully. The electrode pads of the IDE substrate were scraped away using a cotton Q-tip to make electrical connections. The H$_3$PO$_4$ imbibed polycation thin film resistance was determined using electrochemical impedance spectroscopy (EIS). The frequency range was set to 100,000 to 1 Hz with an oscillatory amplitude of 0.1 mA.[47] The equation and IDE dimensions used to calculate in-plane ionic conductivity of the thin films are given in Equation 1 and the descriptions as follows.

$$\kappa = \frac{1}{R}\frac{d}{l\,(N-1)\,t} \qquad \text{(Equation 1)}$$

where

κ: in-plane ionic conductivity
R: in-plane ionic resistance
d: spacing between teeth on IDE (100 μm)
l: length of teeth on IDE (4500 μm)
t: polycation film thickness on IDE substrate
n: number of teeth on IDE substrate (22)

The thickness of the ionomer film was determined by spin coating 1 wt% solution of QPPSf in NMP on silicon wafers (note: identical procedure for IDEs). The silicon wafers were flat and polished and did not contain topographical electrodes. The film thickness was determined via ellipsometry (RC2 Ellipsometer). The thickness values of thin film QPPSf of different IECs ranged from 10–14 nm.

A brief procedure for IEC calculation of pristine QPPSf (i.e., not containing H$_3$PO$_4$) is provided in the Supplemental information.[11] The H$_3$PO$_4$ acid uptake of the QPPSf ionomers synthesized here and targeted as electrode binders were determined by first preparing these materials as binders for gas diffusion electrodes (GDEs). The GDEs were fabricated by spray painting catalyst ink on Toray carbon paper. The procedure for catalyst ink preparation was followed from our previous work.[11] A brief description of the procedure is as follows: 0.2 g of carbon support catalyst (37% Pt in high surface area carbon, Tanaka Kikinzoku International) with 1.715 g of QPPSf ionomer solution dissolved in 5.5 g of reagent alcohol are blended together and sonicated for 30 min to provide a completely dispersed catalyst ink solution. The prepared GDEs were dried at 80°C to remove residual solvent. The catalyst loading was maintained as 0.5 mgPt cm^{-2}. The weight of the non-imbibed GDEs were recorded gravimetrically followed by immersing them in 85 wt% H$_3$PO$_4$ for 10 min for acid imbibing. After the acid-imbibing step, excess liquid acid was carefully removed from the GDE surface via blot drying. Then, the acid-imbibed GDEs were weighed to determine the acid uptake in the electrodes.

Subspace greedy search algorithm

The SGS algorithm seeks the combination of the most contributing subspace of variables that causes the separation between the clusters. To do this, SGS searches from the lowest dimension and moves to higher dimensions until no possible subspaces are left. A testing score, defined as the k-distance between clusters for each subspace, allows for comparison between different subspaces at the same dimension. The subspaces with the highest scores are kept for further testing and the rest are discarded. Further details about the SGS implementation can be found in Zhu et al.[27] This analysis can be performed to explore the relationships between all the clusters and especially the clusters of interest, in this case, the clusters with higher power density. Based on which variables appear more frequently in the contributing subspaces for different clusters, the variable classification can be drawn for the underlying distribution (Supplemental information provides more details about the methodology).

Parameter estimation

When implementing a parameter estimation method, the stochastic or deterministic nature of the method, the complexity of the algorithm, the accuracy, and the convergence speed must be considered. Typically, this will lead to a trade-off between accuracy on one hand and stability, complexity, and convergence speed on the other hand.[47] To address this trade-off, a systematic reduction of the number of parameters to be estimated was enacted by using a global sensitivity analysis. This enhances the performance of a population-based parameter estimation method. The Jaya implementation for our model is described in the Supplemental information and is based off a literature precedent.[48]

AUTHOR CONTRIBUTIONS

Conceptualization, L.A.B., G.V., C.G.A., and J.A.R.; Methodology, L.A.B., G.V., C.G.A., and J.A.R.; Investigation, L.A.B. and G.V.; Data Curation, L.A.B. and J.A.R.; Visualization, L.A.B. and J.A.R.; Experiments, G.V.; Writing – Original Draft, L.A.B., G.V., C.G.A., and J.A.R.; Writing – Review & Editing, L.A.B., G.V., C.G.A., and J.A.R.; Funding Acquisition, C.G.A. and J.A.R.; Resources, C.G.A. and J.R.; Supervision, C.G.A. and J.A.R.

REFERENCES

1. Rosli, R., Sulong, A., Daud, W., Zulkifley, M., Husaini, T., Rosli, M., and Haque, M. (2017). A review of high temperature proton exchange membrane fuel cell (HT-PEMFC) systems. Int. J. Hydrogen Energy *42*, 9293–9314.

2. Xiao, L., Scanlon, E., Ramanathan, L., Choe, E., Rogers, D., Apple, T., and Benicewicz, B. (2005). High-temperature polybenzimidazole fuel cell membranes via a Sol-Gel process. Chem. Mater. *17*, 5328–5333.

3. Lee, K.-S., Maurya, S., Kim, Y., Kreller, C., Wilson, M., Larsen, D., and Mukundan, R. (2018). Intermediate temperature fuel cells via an ion-pair coordinated polymer electrolyte. Energy Environ. Sci. 11, 979–987.

4. Lee, A., Choe, Y.-K., Matanovic, I., and Kim, Y. (2019). The energetics of phosphoric acid interactions reveals a new acid loss mechanism. J. Mater. Chem. A 7, 9867–9876.

5. Quartarone, E., Angioni, S., and Mustarelli, P. (2017). Polymer and composite membranes of proton-conducting, high-temperature fuel cells: a critical review. Materials (Basel) 10, 687.

6. Li, Q., Jensen, J., Savinell, R., and Bjerrum, N. (2009). High temperature proton exchange membranes based on polybenzimidazoles for fuel cells. Prog. Polym. Sci. 34, 449–477.

7. Lee, K.-S., Spendelow, J., Choe, Y.-K., Fujimoto, C., and Kim, Y. (2016). An operationally flexible fuel cell based on quaternary ammonium-biphosphate ion pairs. Nat. Energy 1, 16120.

8. Wainright, J., Wang, J., Weng, D., Savinell, R., and Litt, M. (1995). Acid-doped polybenzimidazoles: a new polymer electrolyte. J. Electrochem. Soc. 142, L121–L123.

9. Pingitore, A., Molleo, M., Schmidt, T., Benicewicz, B., Lipman, T., and Webber, A. (2019). Polybenzimidazole fuel cell technology: theory, performance, and application. In Fuel Cells and Hydrogen Production: A Volume in the Encyclopedia of Sustainability Science and Technology, T.E. Lipman and A. Weber, eds. (Springer), pp. 477–514.

10. Bain, H., Peng, H., Xiang, Y., Zhang, J., Wang, H., Lu, S., and Zhuang, L. (2019). Poly(arylene piperidine)s with phosphoric acid doping as high temperature polymer electrolyte membrane for durable, high-performance fuel cells. J. Power Sources 443, 227219.

11. Venugopalan, G., Chang, K., Nijoka, J., Livingston, S., and Geise, G.M. (2020). Stable and highly conductive polycation-polybenzimidazole membrane blends for intermediate temperature polymer membrane fuel cells. ACS Appl. Energy Mater. 3, 573–585.

12. Chaichi, A., Venugopalan, G., Devireddy, R., Arges, C., and Gartia, M.R. (2020). A solid-state and flexible supercapacitor that operates across a wide temperature range. ACS Appl. Energy Mater. 3, 5693–5704.

13. Kongkanand, A., and Mathias, M. (2016). The priority and challenge of high-power performance of low-Platinum proton-exchange membrane fuel cells. J. Phys. Chem. Lett. 7, 1127–1137.

14. Hu, Y., Jiang, Y., Jensen, J., Cleemann, L., and Li, Q. (2018). Catalyst evaluation for oxygen reduction reaction in concentrated phosphoric acid at elevated temperatures. J. Power Sources 375, 77–81.

15. Mamlouk, M., and Scott, K. (2010). The effect of electrode parameters on performance of a phosphoric acid doped PBI membrane fuel cell. Int. J. Hydrogen Energy 35, 784–793.

16. Correa-Baena, J.-P., Hippalgaonkar, K., van Duren, J., Jaffer, S., Chandrasekhar, V.R., Stevanovic, V., and Bounassi, T. (2018). Accelerating materials development via automation, machine learning, and high-performance computing. Joule 2, 1410–1420.

17. Wang, Y., Seo, B., Wang, B., Zamel, N., Jiao, K., and Cordobes Adroher, X. (2020). Fundamentals, materials, and machine learning of polymer electrolyte membrane fuel cell technology. Energy AI 1, 100014.

18. Ong, S.P. (2019). Accelerating materials science with high-throughput computations and machine learning. Comp. Mater. Sci. 161, 143–150.

19. Blanco, D.E., Lee, B., and Modestino, M.A. (2019). Optimizing organic electrosynthesis through controlled voltage dosing and artificial intelligence. Proc. Natl. Acad. Sci. U S A 116, 17683–17689.

20. Gu, G.H., Noh, J., Kim, I., and Jung, Y. (2019). Machine learning for renewable energy materials. J. Mater. Chem. A 7, 17096–17177.

21. Bao, J., Wang, C., Xu, Z., and Koeppel, B.J. (2019). Physics-informed Machine Learning with Applications to Solid Oxide Fuel Cell System Modeling and Optimization (Pacific Northwest National Lab).

22. Wang, B., Xie, B., Xuan, J., and Kui, J. (2020). AI-based optimization of PEM fuel cell catalyst layers for maximum power density via data-driven surrogate modeling. Energy Convers. Manag. 205, 112460.

23. Cao, B., Adutwum, L.A., Oliynyk, A.O., Luber, E.J., Olsen, B.C., Mar, A., and Buriak, J.M. (2018). How to optimize materials and devices via design of experiments and machine learning: demonstration using organic photovoltaics. ACS Nano 12, 7434–7444.

24. Campello, R.J., Moulavi, D., and Zimek, A. (2015). Hierarchical density estimates for data clustering, visualization, and outlier detection. ACM Trans. Knowl. Discov. Data 10, Article 5.

25. McInnes, L., and Healy, J. (2017). Accelerated Hierarchical Density Clustering. 2017 ICDM Workshops (IEEE).

26. McInnes, L., and Healy, J. (2018). UMAP: uniform manifold approximation and projection for dimension reduction. arXiv, abs/1802.03426.

27. Zhu, W., Sun, W., and Romagnoli, J. (2018). Adaptive k-nearest-neighbor method for process monitoring. Ind. Eng. Chem. Res. 57, 2574–2586.

28. Yin, C., Gao, Y., Li, T., Xie, G., Li, K., and Tang, H. (2020). Study of internal multi-parameter distributions of proton exchange membrane fuel cell with segmented cell device and coupled three-dimensional model. Renew. Energy 147, 650–662.

29. Kregar, A., Tavcar, G., Kravos, A., and Katrasnik, T. (2020). Predictive system-level modeling framework for transient operation and cathode platinum degradation of high temperature proton exchange membrane fuel cells. Appl. Energy 263, 114547.

30. Oh, K., Jeong, G., Cho, E., Kim, W., and Ju, H. (2014). A CO poisoning model for high-temperature proton exchange membrane fuel cells comprising phosphoric acid-doped polybenzimidazole membranes. Int. J. Hydrogen Energy 39, 21915–21926.

31. Sousa, T., Mamlouk, M., and Scott, K. (2010). An isothermal model of a laboratory intermediate temperature fuel cell using PBI phosphoric acid membranes. Chem. Eng. Sci. 65, 2513–2530.

32. Cheddie, D.F., and Munroe, N.D. (2007). A two-phase model of a in intermediate temperature PEM fuel cell. Int. J. Hydrogen Energy 32, 832–841.

33. Scott, K., and Mamlouk, M. (2009). A cell voltage equation for an intermediate temperature proton exchange membrane fuel cell. Int. J. Hydrogen Energy 34, 9195–9202.

34. Drucker, H., Burges, C., Kaufman, L., Smola, A., and Vapnik, V. (1997). Support vector regression machines. Adv. Neural Inf. Process. Syst. 9, 155–161.

35. Boser, B., Guyon, I., and Vapnik, V. (1992). A training algorithm of optimal margin classifiers. In Proceedings of the Fifth COLT, D. Hausler, ed. (ACM Press), pp. 144–152.

36. Abdin, Z., Webb, C.J., and Gray, E. (2016). PEM fuel cell model and simulation in Matlab-Simulink based on physical parameters. Energy 116, 1131–1144.

37. Corrêa, J.M., Farret, F.A., Canha, L.N., and Simões, M.G. (2004). An electrochemical-based fuel-cell model suitable for electrical engineering automation approach. IEEE Trans. Ind. Electron. 51, 1103–1113.

38. Chang, C., and Ling, C. (2011). LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, Article 27.

39. Rao, R.V. (2016). Jaya: a simple and new optimization algorithm for solving constrained and unconstrained optimization problems. Int. J. Ind. Eng. Comput. 7, 19–34.

40. Jeon, Y., Kim, D.J., Koh, J.K., Ji, Y., Kim, J.H., and Shul, Y.-G. (2015). Interface-designed membranes with shape-controlled patterns for high-performance polymer electrolyte membrane fuel cells. Sci. Rep. 5, 16394.

41. Ul Hassan, N., Mandal, M., Huang, G., Firouzjaie, H.A., Kohl, P.A., and Mustain, W.E. (2020). Achieving high-performance and 2000 h stability in anion exchange membrane fuel cells by manipulating ionomer properties and electrode optimization. Adv. Energy Mater. 10, 2001986.

42. Sata, T. (2004). Ion Exchange Membranes (Royal Society of Chemistry).

43. Grew, K.N., Chu, D., and Chiu, W.K.S. (2010). Ionic equilibrium and transport in the alkaline anion exchange membrane. J. Electrochem. Soc. 157, B1024–B1032.

44. Arges, C., Wang, L., Parrondo, J., and Ramani, V. (2013). Best practices for investigating anion exchange membrane suitability for alkaline

electrochemical devices: case study using quaternary ammonium poly (2, 6-dimethyl 1, 4-phenylene) oxide anion exchange membranes. J. Electrochem. Soc. *160*, F1258.

45. Arges, C., and Ramani, V. (2013). Two-dimensional NMR spectroscopy reveals cation-triggered backbone degradation in polysulfone-based anion exchange membranes. Proc. Natl. Acad. Sci. U S A *110*, 2490–2495.

46. Arges, C., Li, K., Zhang, L., Kambe, Y., Wu, G.-P., Lwoya, B., Albert, J.N.L., Nealey, P.F., and Kumar, R. (2019). Ionic conductivity and counterion condensation in nanoconfined polycation and polyanion brushes prepared from block copolymer templates. Mol. Syst. Des. Eng. *4*, 365–378.

47. Su, Z., Kole, S., Harden, L., Palakkal, V., Kim, C., Nair, G., and Renner, J. (2019). Peptide-modified electrode surfaces for promoting anion exchange ionomer microphase separation and ionic conductivity. ACS Mater. Lett. *1*, 467–475.

48. Xu, S., Wang, Y., and Zhi, W. (2019). Parameter estimation of proton exchange membrane fuel cells using eagle strategy based on JAYA algorithm and Nelder-Mead simplex method. Energy *173*, 457–467.